JOURNAL OF MULTIVARIATE ANALYSIS 21, 296–311 (1987)

A Simple Alternative Derivation of a Useful Theorem in Linear "Errors-in-Variables" Regression Models Together with Some Clarifications

YNGVE WILLASSEN

University of Oslo, Oslo, Norway

Communicated by P. R. Krishnaiah

1. INTRODUCTION

The present paper deals with regression theory for linear "errors-in-variables" models. In these models it is well known that the structural vector is not identifiable from the covariance matrix of the observable variables. The modern approach to "solving" this problem is by introducing sufficient prior information to identify the parameters of interest. This prior information may take the form of some a priori knowledge of the covariance matrix of the measurement errors, or specification of a sufficient number of instrumental variables (see the survey papers by Madansky (1959) and Moran (1971)). Although, these approaches are often useful and adequate, we frequently have situations when the additional information needed for identification is simply not available. But, even if the structural vector is not identifiable, we can extract valuable information about it from the covariance matrix of the observable variables. It is this point that is the topic of the present paper.

Frisch (1934) considered the bivariate case and showed that the structural slope parameter was bounded by the two slope parameters obtained from the two regressions. This result was extended to higher dimensions by Reiersøl (1941). For certain conditions on the adjoint of the covariance matrix of the observable variables, Reiersøl (1941, Sects. 4 and 9; 1945, Sects. 9 and 17) proved that the structural vector is contained in a simplex whose vertices are obtained from the different regressions. This means that, although the structural vector cannot be uniquely determined, we know the possible region of its position, which is considerable improvement compared to the situation of complete ignorance. Not until recently (see, e.g.,

Received July 17, 1984; revised March 10, 1985.

AMS 1980 subject classifications: Primary 62H12.

Key words and phrases: errors in variables, regression models.

0047-259X/87 $3.00

Kalman, 1982; Patefield, 1981; Willassen, 1981, 1984), this result has been largely ignored in the literature. The useful survey papers by Madansky and Moran (op.cit.), treat several aspects of "errors-in-variables" models, but only Moran states this result in the bivariate case. One possible explanation of this ignorance is that Reiersøl's approach can be difficult to follow. However, such an appealing result should have a more simple derivation.

In Section 3 we give what we believe to be such a simple derivation. We show this result by using only elementary matrix theory. When the adjoint of the covariance matrix of the observable variables does not satisfy the stated condition, the possible region of the structural vector is unbounded. In this case we briefly indicate how instrumental variables, if available, can be used to bound the possible region.

This point is further illustrated by an example in Section 4. A trivariate example is analysed in this section to show what happens when the stated condition on the covariance matrix fails. When the covariance matrix has incompatible signs, we show that the problem can be decomposed into 3 different cases. Then by an obvious approximation procedure we can use the general theory of Section 3 for each separate case. This decomposition is not restricted to the trivariate case, and it simplifies considerably the treatment of the incompatible sign cases (see Definition 2.4).

For later reference we sketch the model:

$$X_j = \xi_j + \varepsilon_j \qquad j = 1, 2, ..., N \tag{1.1}$$

$$\sum_{j=1}^{N} \gamma_j \xi_j + \gamma_0 = 0. \tag{1.2}$$

In Eqs. (1.1)–(1.2) the observable random variables are denoted by the X's, the unobservable random systematic variables by the ξ's, and finally the random measurement errors by the ε's. The vector $\gamma' = (\gamma_1, \gamma_2, ..., \gamma_N)$ will be called the structural vector. The following assumptions will be used

(i) The set of variables $(\xi_1, \xi_2, ..., \xi_N)$ are distributed in a multivariate distribution with finite first- and second-order moments.

(ii) The set of variables $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)$ are mutually independent with zero expectation and finite variances $(\lambda_1, \lambda_2, ..., \lambda_N)$.

(iii) There is no correlation between the two set of variables $(\xi_1, \xi_2, ..., \xi_N)$ and $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)$.

In Section 3 we use the definitions:

Definition 1.1. (a) M denotes the covariance matrix of the observable X's.

(b) $|M|$ denotes the determinant of M, since we always assume M to be positive definite $|M| > 0$.

(c) $|M_{ij}|$ denotes the cofactor of the (ij) element of M.

DEFINITION 1.2. L will denote a diagonal matrix with non-negative elements $\lambda_1, \lambda_2, \dots, \lambda_N$.

DEFINITION 1.3. $(M - L)$ will denote the covariance-matrix of $(\xi_1, \xi_2, \dots, \xi_N)$ when the diagonal elements of L are interpreted as variances of $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)$.

2. DEDUCTION OF SOME USEFUL MATHEMATICAL RESULTS

In the proofs of our main theorems (Sect. 3) we will need a few special results in matrix theory. For the sake of clarity in presentation we shall list the necessary definitions and prove these auxiliary theorems in the present section.

DEFINITION 2.1. For a given matrix A:

(i) $|A|$ denotes the determinant of A;

(ii) $|A_{ij}|$ denotes the cofactor of the (ij)th element of A. When $i = j$ we call $|A_{ij}|$ a principal minor.

DEFINITION 2.2. For a given symmetric and positive definite matrix A, $\mathscr{L}$ denotes the set of diagonal matrices $L = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_G)$ with non-negative elements such that $L \in \mathscr{L}$ implies:

(i) $(A - L)$ is nonnegative definite;

(ii) $|A - L| = 0$.

DEFINITION 2.3. A diagonal matrix T is called a sign matrix if any of its diagonal elements are 1 or -1.

DEFINITION 2.4. (i) A matrix A is said to have compatible signs or it is called compatible if there exists a sign matrix T such that (TAT) has positive elements only.

(ii) If there exists a sign matrix T such that the elements of (TAT) are nonnegative but at least one element is zero, then A is said to have semi-compatible signs.

(iii) If (TAT) contains both positive and negative elements for any sign matrix T, then A is said to have incompatible signs.

THEOREM 2.1. *Let A be a $G \times G$ matrix with negative off-diagonal elements and nonnegative principal minors of all orders. Then, the rank of A is at least $(G-1)$ and the adjoint of A has positive elements only.*

Proof. Let B denote the adjoint of A, $B = \operatorname{adj} A$. If $G=2$ $a_{11}a_{22} - a_{12}a_{21} \geqslant 0$ together with $a_{12} < 0$ and $a_{21} < 0$ imply $a_{11} > 0$ and $a_{22} > 0$. Hence, Theorem 2.1 is true for $G=2$. Let us partion A and B as

$$A = \begin{pmatrix} A_{G-1} & a_{G1} \\ a_{1G} & a_{GG} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{G-1} & b_{G1} \\ b_{1G} & b_{GG} \end{pmatrix}.$$

Suppose that the theorem is true for the $(G-1) \times (G-1)$ submatrix A_{G-1} obtained from A by deleting the last row and column. By the identity (11.5.3) (Cramér, 1946, p. 109) and $|A| \geqslant 0$,

$$a_{GG}|A_{G-1}| - a_{1G}(\operatorname{adj} A_{G-1})\, a_{G1} = |A| \geqslant 0. \tag{2.1}$$

By induction, (adj A_{G-1}) has positive elements and, by hypothesis, a_{1G} and a_{G1} have negative elements $a_{GG} \geqslant 0$ and $|A_{G-1}| \geqslant 0$. Then (2.1) implies

$$a_{GG}|A_{G-1}| \geqslant a_{1G}(\operatorname{adj} A_{G-1})\, a_{G1} > 0, \tag{2.2}$$

implying $a_{GG} > 0$ and $|A_{G-1}| > 0$. Hence, the rank of A is at least $G-1$. Note that $b_{GG} = |A_{G-1}| > 0$ and $AB = BA = |A|\, I$ give after some simplifications

$$\begin{aligned} b_{G1} &= -A_{G-1}^{-1} a_{G1} b_{GG} = -(\operatorname{adj} A_{G-1})\, a_{G1} \\ b_{1G} &= -a_{1G}(\operatorname{adj} A_{G-1}) \\ |A_{G-1}|\, B_{G-1} &= (\operatorname{adj} A_{G-1})|A| - b_{G1} a_{1G}(\operatorname{adj} A_{G-1}) \\ &= (\operatorname{adj} A_{G-1})|A| + b_{G1} b_{1G}. \end{aligned}$$

Now, arguing as above, we find that b_{G1}, b_{1G}, and B_{G-1} have only positive elements. Q.E.D.

THEOREM 2.2. *Let A be a $G \times G$ positive definite symmetric matrix and let $L \in \mathscr{L}$. Suppose that one of the co-factors $|A_{kj}|$ $(k, j = 1, 2, \ldots, G;\ j \neq k)$ is zero. Then there exists a diagonal matrix $L \in \mathscr{L}$ such that the principal minor $|B_{kk}(L)|$ of the matrix $B(L) = (A - L)$ is equal to zero.*

Proof. Suppose $|A_{kj}| = 0$ for some k and j $(k \neq j)$. Then, by the identity (11.7.3) (Cramér, 1946, p. 111)

$$|A|\ |A_{kk \cdot jj}| = |A_{kk}|\ |A_{jj}| - |A_{kj}|^2, \tag{2.3}$$

where $|A_{kk \cdot jj}|$ denotes the principal minor obtained from A by deleting kth and jth rows and columns. Choose $L \in \mathscr{L}$ such that $\lambda_i = 0$ for all $i \neq j$ and $\lambda_j = |A|/|A_{jj}|$. Then, notice that $(A - L)$ is nonnegative definite, $|A - L| = |A| - \lambda_j |A_{jj}| = 0$, and

$$|B_{kk}(L)| = |A_{kk}| - \lambda_j |A_{kk \cdot jj}| = |A_{kj}|^2 / |A_{jj}| = 0. \qquad \text{Q.E.D.}$$

Since, $L = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_G)$ we can interprete L as a point in a G-dimensional Euclidean space.

THEOREM 2.3. *Let A be a $G \times G$ symmetric positive definite matrix. Suppose there exists a point (matrix) $L_* \in \mathscr{L}$, where a principal minor of $B(L_*) = (A - L_*)$, for example, $|B_{11}(L_*)|$ vanishes. Then there exists a point $L_{**} \in \mathscr{L}$ on the straight line through L_* and parallel to the λ_1 axis, where the eigenvalue $\Theta = 0$ of the characteristic equation $|B(L_{**}) - \Theta I| = 0$ has multiplicity 2 at least.*

Proof. Since $L_* \in \mathscr{L}$ the matrix $B(L_*)$ is nonnegative definite and $|B(L_*)| = 0$. This fact together with the hypothesis $|B_{11}(L_*)| = 0$ implies that the minors $|B_{12}(L_*)|, \ldots, |B_{1G}(L_*)|$ also vanish (use identity (11.7.3) (Cramér, op.cit.). Denote by L_1 the diagonal matrix whose elements are all equal to those of L_*, except for the first element λ_1 which is now permitted to be varied freely. Evidently by varying λ_1 we can interprete L_1 as points on a straight line in a G-dimensional space. For an arbitrary value of λ_1 put $B(L_1) = (A - L_1)$. Since the minors $|B_{11}(L_1)|, |B_{12}(L_1)|, \ldots, |B_{1G}(L_1)|$ are independent of λ_1, and since L_1 is equal to L_* except for the first element λ_1, it follows that $|B_{11}(L_1)|, |B_{12}(L_1)|, \ldots, |B_{1G}(L_1)|$ all vanish. This implies that $|B(L_1)| = 0$ for any value of λ_1. Hence, $\Theta = 0$ is a root of $|B(L_1) - \Theta I| = 0$ for any value of λ_1.

Let λ_1^0 be a fixed value of λ_1 and let L_1^0 denote the corresponding diagonal matrix. If $B(L_1^0) = (A - L_1^0)$ is nonnegative definite it is easy to see that $B(L_1)$ is nonnegative definite for any value of λ_1 in the interval $[0, \lambda_1^0]$. Hence, for sufficiently small values of λ_1, the matrix $B(L_1)$ is nonnegative definite; but by increasing λ_1 we can make $B(L_1)$ indefinite. That is, if we choose λ_1 sufficiently large, at least one of the roots of $|B(L_1) - \Theta I| = 0$, for example, $\Theta_1(\lambda_1)$ is negative. However, $\Theta_1(\lambda_1)$, being a root of $|B(L_1) - \Theta I)| = 0$, is a continuous function with respect to λ_1. Since $\Theta_1(\lambda_1) \geqslant 0$ for sufficiently small values of λ_1, it follows that by increasing λ_1 there exists a point $\lambda_1 = \lambda_1^{**}$, where $\Theta_1(\lambda_1) = 0$. Hence, if L_{**} denotes the diagonal matrix whose elements are equal to those of L_*, except for the first element which is equal to λ_1^{**}, then $\Theta = 0$ is a root of $|B(L_{**}) - \Theta I| = 0$ of multiplicity 2 at least. Q.E.D.

As a consequence of this theorem we can state

THEOREM 2.4. *Let A and L_* be the matrices defined in Theorem 2.3. Suppose that a principal minor, for example, $|B_{11}(L_*)|$ vanishes. Then we can find a diagonal matrix $L_{**} \in \mathscr{L}$ such that the rank of $B(L_{**}) = (A - L_{**})$ is not larger than $G-2$.*

THEOREM 2.5. *Let A be a $G \times G$ symmetric positive definite matrix. Then* adj A *has compatible signs if and only if for all $L \in \mathscr{L}$ the elements of* $\operatorname{adj}(A-L)$ *are nonzero.*

Proof. Suppose adj A has compatible signs and let T be the sign matrix such that $E^{-1} = (TA^{-1}T)$ has positive elements only. Since $E = (TAT)$, we have

$$|A - L| = 0 \Leftrightarrow |E - L| = 0 \Leftrightarrow |(I - E^{-1}L)| = 0 \tag{2.4}$$

(remember $T^2 = I$, $T^{-1} = T$, adj $T = \pm T$).

We notice that all principal minors of $(I - E^{-1}L)$ are nonnegative. If $L = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_G) \in \mathscr{L}$ is non-singular, all off-diagonal elements of $(I - E^{-1}L)$ are negative. By Theorem 2.1 the matrix $\operatorname{adj}(I - E^{-1}L)$ has then only positive elements. Since,

$$\operatorname{adj}(E - L) = \operatorname{adj}(I - E^{-1}L)(\operatorname{adj} E) \tag{2.5}$$

the elements of $\operatorname{adj}(E-L)$ are positive when the elements of (adj E) are positive. Consequently, since

$$T(\operatorname{adj}(A - L))\, T = \operatorname{adj}(E - L), \tag{2.6}$$

it follows that all elements of $\operatorname{adj}(A-L)$ are nonzero when adj A has compatible signs. It is not difficult to show that the same conclusion holds when $L \in \mathscr{L}$ is singular.

Conversely, suppose that $\operatorname{adj}(A-L)$ has nonzero elements for every $L \in \mathscr{L}$. Since the elements of $\operatorname{adj}(A-L)$ are continuous functions of the elements of L, the elements of $\operatorname{adj}(A-L)$ preserve the same signs for all $L \in \mathscr{L}$. Since $(A-L)$ is symmetric and singular the rows of $\operatorname{adj}(A-L)$ are proportional, and $\operatorname{adj}(A-L)$ will have compatible signs. Consider the matrix $L \in \mathscr{L}$ when all of its elements are zero, except the jth element (i.e., $\lambda_j = |A|/A_{jj}|$). At this point the jth row of $\operatorname{adj}(A-L)$ coincides with the jth row of adj A. Then it follows that adj A has compatible signs. Q.E.D.

Further, since $|A - L| = 0$ for all $L \in \mathscr{L}$, we note the following facts by applying the identity (2.3) to the matrix $(A-L)$. If an off-diagonal element of $\operatorname{adj}(A-L)$ is zero, then at least one diagonal element of $\operatorname{adj}(A-L)$ is zero. Hence, if an element of $\operatorname{adj}(A-L)$ is zero, then at least one principal minor of $(A-L)$ is zero. This is what we need in Theorem 2.3.

Finally, we notice that Theorems 2.3 and 2.5 tell us that it is the Perron version of the Perron–Frobenius theorem we shall use in our proof of our main theorem in the next section.

3. The Mapping from $L \in \mathscr{L}$ to $\gamma \in \Gamma$ Determined by the Equation $(M-L)\gamma = 0$.

If we multiply Eq. (1.2) by $(\xi_j - E\xi_j)$ $(j = 1, 2, \ldots, N)$ and follow the assumptions (i)–(iii), we obtain the matrix equation

$$(M-L)\gamma = 0 \quad M,\ L \in \mathscr{L} \text{ as defined in Section 1.} \tag{3.1}$$

The following theorems will hold if M^{-1} has compatible signs. That is, if there exists a sign matrix T such that $(TM^{-1}T)$ has only positive elements. Noting this fact and simplifying notation, we use the condition: M^{-1} has only positive elements. Finally, if $L \in \mathscr{L}$ is singular the necessary modifications of the theorems are obvious and straightforward. We therefore suppose that $L \in \mathscr{L}$ is non-singular, and afterwards indicate the necessary modifications for the case when L is singular.

THEOREM 3.1. *Let M^{-1} be a symmetric and positive definite matrix with ony positive elements. Suppose that the diagonal matrix $L \in \mathscr{L}$ is non-singular. Then, $(M-L)\gamma = 0$ determines the structural vector γ uniquely except for a scalar factor, and γ is proportional to a positive vector.*

Proof (Reiersøl, 1941, pp. 8–9, using the Perron–Frobenius theorem). Hence, under the hypotheses of Theorem 3.1, we can always normalize the structural vector γ by $(\gamma)\gamma_1^{-1} = \binom{1}{g}$ and then $(M-L)\gamma = 0$ is equivalent to $(M-L)\binom{1}{g} = 0$.

We have now the necessary technical background to show the main theorems in confluence analysis. The following definitions will be used:

e'—the row vector whose elements are all 1.

D—the diagonal matrix with elements consisting of the first row of M^{-1}.

$\binom{e'}{P} = (M^{-1}D^{-1})$, where P consists of $N-1$ rows and N columns to be denoted $P_1, P_2, \ldots, P_N$.

$\mathscr{P}$—the $(N-1)$-dimensional simplex generated by the columns (points) $P_1, P_2, \ldots, P_N$.

THEOREM 3.2. *Suppose that M^{-1} and $L \in \mathscr{L}$ satisfy the hypotheses of Theorem 3.1. Then, $(M-L)\binom{1}{g} = 0$ determines a mapping from $\mathscr{L}$ to the* *simplex $\mathscr{P}$, such that to every $L \in \mathscr{L}$ there corresponds one and only one $g \in \mathscr{P}$, and to every $g^* \in \mathscr{P}$ there corresponds one and only one $L^* \in \mathscr{L}$.*

Proof. Suppose $L \in \mathscr{L}$. $(M-L)\binom{1}{g}=0 \Leftrightarrow (I-LM^{-1})\,M\binom{1}{g}=0$. Put $c=M\binom{1}{g}$ and notice that according to Theorem 3.1 the elements of c are positive. Further,

$$\begin{pmatrix}1\\ g\end{pmatrix}=M^{-1}c=(M^{-1}D^{-1})\,Dc=\begin{pmatrix}e'\\ P\end{pmatrix}w, \qquad w=Dc \tag{3.2}$$

Since the elements of the diagonal matrix D are positive, the elements of the column vector $w=Dc$ are also positive. From (3.2) it follows that

$$1=e'w=\sum_{i=1}^{N} w_i \tag{3.3a}$$

$$g=Pw=\sum_{i=1}^{N} w_i P_i. \tag{3.3b}$$

Equations (3.3a)–(3.3b) show that g is contained in the simplex $\mathscr{P}$; and to a given $L \in \mathscr{L}$ there corresponds one g.

Conversely, suppose that $g^* \in \mathscr{P}$. Then there exists a set of nonnegative numbers $(v_1, v_2, \ldots, v_N)$ added to 1 so that $g^*=\sum_{i=1}^{N} v_i P_i$. This is equivalent to

$$\begin{pmatrix}1\\ g^*\end{pmatrix}=\begin{pmatrix}e'\\ P\end{pmatrix}v \qquad (v'=(v_1, v_2, \ldots, v_N)). \tag{3.4}$$

Put $c^*=D^{-1}v$; the elements of c^* are certainly nonnegative. Equation (3.4) then implies

$$\begin{pmatrix}1\\ g^*\end{pmatrix}=\begin{pmatrix}e'\\ P\end{pmatrix}Dc^*=(M^{-1}D^{-1})\,Dc^*=M^{-1}c^*. \tag{3.5}$$

Since, by hypotheses, the elements of M^{-1} are positive, the elements of g^* determined by (3.5) are also positive. Define the diagonal matrix L^* by

$$c^*=L^*\begin{pmatrix}1\\ g^*\end{pmatrix}. \tag{3.6}$$

The elements of L^* are evidently nonnegative. (An element of L^* is zero if the corresponding element of c^* is zero). Finally, Eqs. (3.5) and (3.6) taken together imply

$$(M-L^*)\begin{pmatrix}1\\ g^*\end{pmatrix}=0. \tag{3.7}$$

Since $\binom{1}{g^*}$ is a positive vector, the Perron–Frobenius theorem implies that $L^* \in \mathscr{L}$. Q.E.D.

We note that the points $P_1, P_2, \ldots, P_N$ generating the simplex $\mathscr{P}$ are just the vectors obtained by the N regressions, that is, the regressions of X_1 on $X_2, \ldots, X_N$; the regression of X_2 on $X_1, X_3, \ldots, X_N$; and then X_1 in terms of $X_2, X_3, \ldots, X_N$, etc. Our arguments above also show that if $L \in \mathscr{L}$ is singular, i.e., $L = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_J, 0, 0, \ldots, 0)$ then the simplex $\mathscr{P}$ is generated by the points $P_1, \ldots, P_J$.

The condition that M^{-1} has compatible signs is also necessary as shown by the following two results.

THEOREM 3.3. *Suppose that M^{-1} has incompatible signs. Then there exists a diagonal matrix $L_{**} \in \mathscr{L}$ such that the solutions of $(M - L_{**})\,\gamma = 0$ constitute a linear manifold of dimension* 2 *at least.*

Proof. Suppose M^{-1} has incompatible signs. Then, according to Theorem 2.5, we can find a matrix $L_* \in \mathscr{L}$ so that at least one element of $\mathrm{adj}(M - L_*)$ is zero. If an off-diagonal element of $\mathrm{adj}(M - L_*)$ is zero, we can, by the identity (2.3) applied to $(M - L_*)$, find at least one diagonal element of $\mathrm{adj}(M - L_*)$ which is zero since $|M - L_*| = 0$. From Theorem 2.3 we then know that there exists a matrix $L_{**} \in \mathscr{L}$ such that the rank of $(M - L_{**})$ is at most $(N - 2)$. The conclusion then follows immediately. Q.E.D.

Consider Eq. (1.2) and suppose that $\gamma_1 \neq 0$. Then we can put $\binom{1}{g} = (\gamma)\,\gamma_1^{-1}$ and $(M - L)\,\gamma = 0$ becomes equivalent to $(M - L)\binom{1}{g} = 0$.

THEOREM 3.4. *Suppose that M^{-1} has incompatible signs and suppose $\gamma_1 \neq 0$. Then there exists a diagonal matrix $L_{**} \in \mathscr{L}$ such that the general solution of $(M - L_{**})\binom{1}{g} = 0$ has at least one degree of freedom.*

Proof. Use the same arguments as in the proof of the preceding theorem. Q.E.D.

Hence, the possible range of each of the coefficients $g_2, \ldots, g_N$ are from $-\infty$ to $+\infty$ by proper choice of the free parameter(s). Therefore, we can make g outside any bounded set, and certainly outside the simplex $\mathscr{P}$.

Above we assumed that $L \in \mathscr{L}$ is non-singular or that there are noise (measurement errors) in all N observable variables. But what happens if we know a priori that only some of the variables are observed with noise. Without loss of generality we assume that $X_1, X_2, \ldots, X_J$ are observed with noise ($1 \leqslant J < N$), and the remaining $K = N - J$ variables are observed exactly. Then we partition the matrices M, M^{-1}, and L correspondingly

$$M = \begin{pmatrix} M_{JJ} & M_{JK} \\ M_{KJ} & M_{KK} \end{pmatrix}, \qquad M^{-1} = \begin{pmatrix} M^{JJ} & M^{JK} \\ M^{KJ} & M^{KK} \end{pmatrix} \tag{3.8a,b}$$

$$L = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_J, 0, ..., 0) = \begin{pmatrix} L_J & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.8c}$$

where L_J is a positive definite diagonal matrix.

As above our main concern is the rank of $(M - L_J)$ when $L_J \in \mathscr{L}$. First, we note that

$$|(M - L_J)| = 0 \Leftrightarrow |(I - L_J M^{-1})| = 0 \tag{3.9a,b}$$

Using the partitions (3.8a)–(3.8c) the matrix $(I - L_J M^{-1})$ can be written

$$G = \begin{pmatrix} (I_J - L_J M^{JJ}) & -L_J M^{JK} \\ 0 & I_K \end{pmatrix}, \tag{3.10}$$

where I_J and I_K denote identity matrices. Hence, (3.9b) implies

$$|I_J - L_J M^{JJ}| = 0. \tag{3.11}$$

We note that

$$M^{JJ} = (M_{JJ} - M_{JK} M_{KK}^{-1} M_{KJ})^{-1} \tag{3.12}$$

so that (3.11) implies

$$|(M_{JJ} - M_{JK} M_{KK}^{-1} M_{KJ}) - L_J| = 0. \tag{3.13}$$

Since M is positive definite the matrix $(M_{JJ} - M_{JK} M_{KK}^{-1} M_{KJ})$ is also positive definite (see, e.g., Muirhead, 1982, p. 586). Hence, we can apply the results obtained in Section 2 to this matrix. Consequently, we observe that the rank of $(M - L_J)$ is essentially dependent on the signs of the elements of the submatrix M^{JJ}. If M^{JJ} has compatible signs the rank of $(I_J - L_J M^{JJ})$ is $(J-1)$ for all $L_J \in \mathscr{L}$, and from (3.10) we then observe that the rank of $(M - L_J)$ is $J - 1 + K = N - 1$ for all $L_J \in \mathscr{L}$. This conclusion holds even if the other submatrices of M^{-1} have incompatible signs. If M^{JJ} has incompatible signs then we can find a $L_J \in \mathscr{L}$ so that the rank of $(I_J - L_J M^{JJ})$ is at most $(J-2)$. By (3.10) the rank of $(M - L_J)$ is at most $(N-2)$, and therefore, the solution space of $(M - L_J)\gamma = 0$ is at least of dimension 2 at this $L_J \in \mathscr{L}$.

Then, let us finally sum up this situation. If $J=1$ there is noise in one variable only, e.g., X_1. The rank of L_J is one, the matrix (3.12) reduces to the positive element $|M_{11}|/|M|$ (see the identity 2.1). The "submatrix" M^{JJ} has therefore compatible sign in this case, and according to what has been said above the solution of $(M-L_J)\gamma=0$ or

$$(I-L_JM^{-1})g=0 \qquad g=M\gamma \tag{3.14}$$

is unique (except for a constant). From (3.11) we observe that the element of L_1 is equal to $|M|/|M_{11}|$. Normalizing so that $\gamma_1=1$ we obtain the least-squares solution.

If $J=2$ we observe two variables with noise, e.g., X_1 and X_2. The submatrix M^{JJ} is then (2×2) and this matrix has always compatible signs except in the case the cofactor $|M_{12}|=0$. In the latter case we observe from (3.11) and (3.10) that if $L_2=\operatorname{diag}(|M|/|M_{11}|,\ |M|/|M_{22}|)\in\mathscr{L}$ the rank of $(M-L_2)$ or $(I-L_2M^{-1})$ is $N-2$.

This discussion should be sufficient to show what happens when only some of the variables are observed with noise.

Finally, Theorems (3.3) and (3.4) seem rather detrimental from a statistical point of view. These results may, however, be useful indeed, if the model determines one or more other relationships between the structural coefficients. In particular applications the models will often specify a set of auxiliary variables which are suitable as instrumental variables (see Moran, 1971, Sect. 12). Suppose, for instance, that the application of an instrumental variable gives a linear equation between the coefficients $g_2,..., g_N$. Then the possible region of the structural vector g will be restricted to the intersection between the region which follows according to Theorem 3.4 and the hyperplane H which represents the linear equation mentioned . In the example of the next section we shall illustrate this idea.

4. An Example in the Trivariate Case When (adj M) Does not Have Compatible Signs

Let us regard $L=\operatorname{diag}(\lambda_1, \lambda_2,..., \lambda_N)$ as a point in a N-dimensional space. Then for a given covariance matrix M the set of points $\mathscr{L}$ of definition (2.2) constitutes that part of the surface $|M-L|=0$ which is situated nearest to the origin in the positive orthant (Reiersøl, 1941, Sect. 8).

We consider the example

$$M=\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}. \tag{4.1}$$

The adjoint of this covariance matrix has incompatible signs. By Theorems 2.5 and 2.3 we can find a matrix $L = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) \in \mathscr{L}$ at which the rank of $(M - L)$ is 1. The coordinates of this point can be determined by

$$|(M-L)_{12}| = 0, \quad |(M-L)_{13}| = 0, \quad \text{and} \quad |(M-L)_{23}| = 0. \tag{4.2a)-(4.2c}$$

We obtain

$$\lambda_1^0 = \lambda_2^0 = \lambda_3^0 = 1. \tag{4.3a)-(4.3c}$$

We also observe that the (point) matrix $L_0 = \mathrm{diag}(\lambda_1^0, \lambda_2^0, \lambda_3^0)$ is uniquely determined by (4.2a)–(4.2c), and notice that the rank of $(M - L_0)$ is 1.

Hence, all principal minors of order 2 will also vanish at this point. Let $(L_0 B)_i$ denote the straight line segment parallel to the λ_i axis with end points L_0 and B_i. B_i denotes the point where this line segment intersects the coordinate plane $\lambda_i = 0$. Since $|(M-L)_{ii}|$ is independent of λ_i, $|(M-L)_{ii}|$ vanishes at every point of the segment $(L_0 B)_i$. Similarly we observe that the determinant $|M - L|$ vanishes when L is any point of the line segment $(L_0 B)_i$ $(i = 1, 2, 3)$. We conclude that these three straight line segments partition the set of points $\mathscr{L}$ into 3 disjoint regions $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}_3$.

Then we shall find the image in the Γ-space of the set $\mathscr{L}$, determined by

$$\begin{pmatrix} 2-\lambda_1 & -1 & -1 \\ -1 & 2-\lambda_2 & 1 \\ -1 & 1 & 2-\lambda_3 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} = 0, \qquad L = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) \in \mathscr{L}. \tag{4.4}$$

We notice that along the segment $(L_0 B)_1$, we obtain from (4.4)

$$(1-\lambda_1)\gamma_1 = 0 \qquad \text{implying } \gamma_1 = 0 \quad \text{for } 0 \leqslant \lambda_1 < 1. \tag{4.5}$$

Hence, along this segment we cannot normalize by putting $g_i = \gamma_i/\gamma_1$ $(i = 1, 2, 3)$. However, apart from this segment the above normalization rule is allowed for any $L \in \mathscr{L}$.

The following facts are easily observed from (4.4). The point $L_0 = \mathrm{diag}(1, 1, 1)$ is mapped into the straight line

$$g_2 + g_3 = 1. \tag{4.6}$$

The segments $(L_0 B)_2$ and $(L_0 B)_3$ are mapped into the points (0, 1) and (1, 0). Then we shall determine the images of $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}_3$ by direct use of the results of the preceding section.

$\mathscr{L}_1$ is that part of $\mathscr{L}$ where $\lambda_1 \geqslant \lambda_1^0 = 1$, $\lambda_2 \geqslant 0$, $\lambda_3 \geqslant 0$. Suppose that δ is a small positive number. We realize that we can determine the image of $\mathscr{L}_1$ by applying Theorem 3.2 to

$$\begin{pmatrix} 2-(\lambda_1^0+\delta) & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1-\delta & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}, \tag{4.8}$$

since this matrix has compatible signs, and then let δ tend to zero. If $L_0 = \mathrm{diag}(1, 1, 1)$ is considered to be a point of $\mathscr{L}_1$, we conclude, by letting $\delta \to 0$, that $\mathscr{L}_1$ is mapped into the triangle generated by the regression vectors: $P_1 = (\frac{1}{3}, \frac{1}{3})$, $P_2 = (1, 0)$, $P_3 = (0, 1)$ and the straight line through $(1, 0)$ and $(0, 1)$.

$\mathscr{L}_2$ is that part of $\mathscr{L}$ where $\lambda_1 \geqslant 0$, $\lambda_2 \geqslant \lambda_2^0 = 1$, $\lambda_3 \geqslant 0$. By similar analysis applied to the matrix

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2-(\lambda_2^0+\delta) & 1 \\ -1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & (1-\delta) & 1 \\ -1 & 1 & 2 \end{pmatrix}, \tag{4.9}$$

we conclude that the structural vector is contained in a triangle with vertices: $P_1 = (1/(1-2\delta), -\delta/(1-2\delta))$; $P_2 = (3, -1)$; $P_3 = (1/\delta, -(1-2\delta)/\delta)$. We notice that for any δ, P_3 is a point on the straight line

$$g_2 + g_3 = 2. \tag{4.10}$$

By letting $\delta \to 0$ we conclude that $\mathscr{L}_2$ is mapped into a "triangle" with vertices $P_1 = (1, 0)$, $P_2 = (3, -1)$, and P_3, where P_3 moves infinitely far out in the southeast direction on the line (4.10) as $\delta \to 0$.

$\mathscr{L}_3$ is that part of $\mathscr{L}$ where $\lambda_1 \geqslant 0$, $\lambda_2 \geqslant 0$, $\lambda_3 \geqslant \lambda_3^0 = 1$. By applying the arguments above to the matrix

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 1-\delta \end{pmatrix}, \tag{4.11}$$

we conclude that $\mathscr{L}_3$ is mapped into the "triangle" with vertices $P_1 = (0, 1)$, P_2, $P_3 = (-1, 3)$, where P_2 moves infinitely far out in the northwest direction on the line (4.10) as $\delta \to 0$.

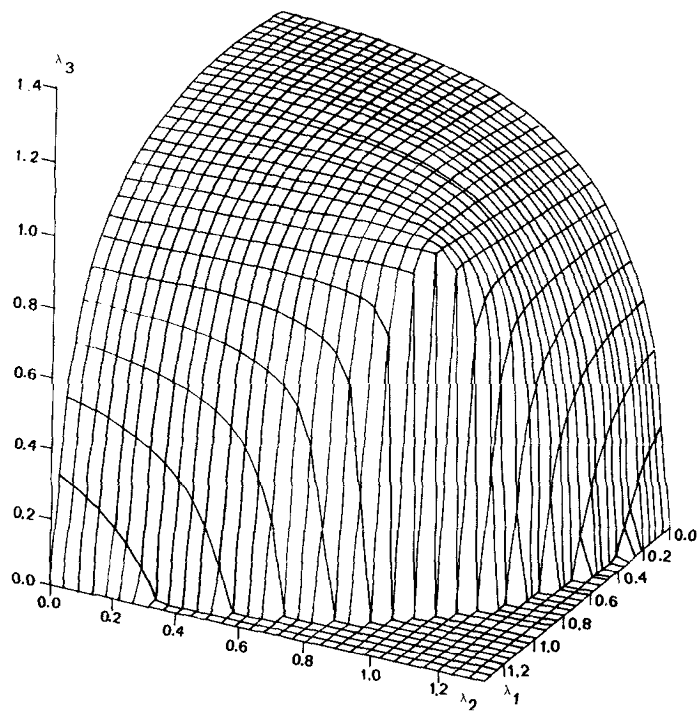
The above trivariate example illustrates the typical findings when the adjoint of the covariance matrix M has incompatible signs. In particular we notice that the partition of $\mathscr{L}$ into 3 disjoint parts makes it possible for us to have 3 corresponding modifications of the original matrix.

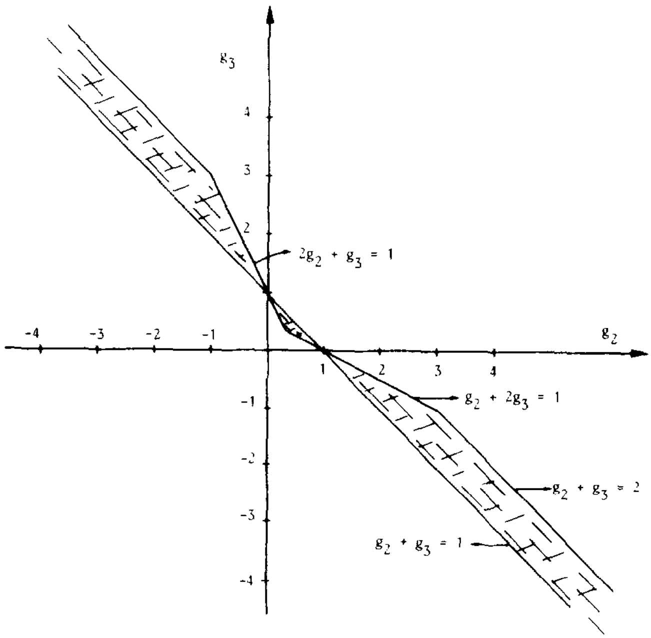
Hence, our study of a covariance matrix with incompatible signs is reduced to 3 separate studies in which the modified covariance matrices have compatible or semi-compatible signs. To these cases we could apply the general theory of Section 3 quite directly. In our opinion this reduction shows that the "incompatible signs case" is not as bad a case as some researchers believe it to be. Although, we do not know $L = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ exactly, we will often know to which part of $\mathscr{L}$ L belongs. But then we have considerable knowledge of the position of the structural vector, too.

Finally, consider the example above and assume that we can normalize $g_i = \gamma_i/\gamma_1$ $(i = 1, 2, 3)$. Suppose we also observe an instrumental variable Z (see Moran, 1971, Sect. 12). Then we obtain

$$m_{1Z} + g_2 m_{2Z} + g_3 m_{3Z} = 0 \qquad (m_{iZ} = \mathrm{Cov}(X_i, Z)). \tag{4.12}$$

Hence by using (4.12) we realize that the possible region of the structural vector is now a bounded set unless (4.12) is parallel to (4.10) and lies between (4.10) and (4.6). A geometrical picture of this example is given.





5. Conclusion

Our motives for writing this paper have been twofold. First, the main theorem (Theorem 3.2) is not properly understood and referred to in the present growing literature in this field. And truly, some parts of Reiersøl's original proof can be difficult to follow. A simple proof of this theorem using only standard matrix theory seemed to be needed. This is given in Section 3 where the problem posed in this section is shown to have a clear structure in the compatible sign case. Second, for further clarification we consider a trivariate example where the condition on the adjoint of the covariance matrix fails. A decomposition of the problem then appears which makes it possible to apply Theorem 3.2 repeatedly to fully analyse the incompatible sign case. Hence, also in this case the problem posed in Section 3 has a clear structure. This unification of the two cases is interesting, and should also work in higher dimensions.

Acknowledgments

I have benefitted from several discussions with Professor O. Reiersøl. I am also thankful to two referees for several useful comments which made it possible to shorten and improve an earlier version of this paper.

REFERENCES

[1] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press, Princeton, N.J.

[2] FRISCH, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems.* Publ. Univ. Inst. of Economics, Oslo.

[3] KALMAN, R. E. (1982). System identification from noisy data. In *Proceedings, International Symposium on Dynamical Systems* (A. Bednarek, Ed.). Academic Press, New York/London.

[4] MADANSKY, A. (1959). The fitting of straight lines when both variables are subject to errors. *J. Amer. Statist. Assoc.* 173–205.

[5] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiely, New York.

[6] MORAN, P. A. P. (1971). Estimating structural and functional relationships. *J. Multivar. Anal.* **1** 232–255.

[7] PATEFIELD, W. M. (1981). Multivariate linear relationships: maximum likelihood estimation and regression bounds. *J. R. Statist. Soc. Ser. B* **43** 342–352.

[8] REIERSØL, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis, *Econometrica*, 1–24.

[9] REIERSØL, O. (1945). Confluence analysis by means of instrumental sets of variables. *Ark. Mat. Astronom. Fys.* K. Svenska Vetenskapsakademien. [Swedish]

[10] WILLASSEN, Y. (1981). Bounds on the structural vector in two mixed linear "errors-in-variables" regression models. unpublished.

[11] WILLASSEN, Y. (1984). Testing hypotheses on the unidentifiable structural parameters in the classical "errors-in-variables" with application to Friedman's permanent income model. *Econom. Lett.* 221–228.